APPENDIX C:
PENDING CLAIMS
UPON ENTRY OF THE PRELIMINARY AMENDMENT

U.S. PATENT APPLICATION SERIAL NO. 09/644,937
(ATTORNEY DOCKET NO. 9476-003-999)

1. (Amended) A computer-implemented method of finding, in a group of N objects, those objects whose minimal metric distance from a first object is less than a threshold distance, X, comprising:

selecting a number M of the N objects, wherein M is much less than N and wherein M is a dimensionality of a shape space of the group of objects and wherein said number M of objects represents said shape space;

making an ordered list of minimal metric distances between each of the M objects and each of the other N objects;

determining the minimal metric distances between the first object and each of the M objects, thereby identifying a second object of said M objects that has the smallest minimal metric distance between itself and the first object;

calculating a minimal metric distance between the first object and at least one object on the ordered list associated with said second object, by:

beginning with the object on said ordered list that has the smallest minimal metric distance between it and said second object and continuing with objects having increasingly greater minimal metric distances from said second object until an object is reached that has a minimal metric distance from said second object that is more than twice the minimal metric distance from said first object to said second object, or whose minimal metric distance from said second object is more than twice the threshold distance from the first object;

repeating said calculating step wherein said second object has a next smallest minimal metric distance from the first object until each of said M objects has been considered; and

selecting those objects from said calculating step whose minimal metric distance from the first object is less than X.

2. (New) The method of claim 1 wherein each of said objects is represented by a field.

3. (New) The method of claim 2 wherein said first object is a vacant space in a pocket on a surface of a molecule and is represented by a field.

4. (New) The method of claim 2 wherein each of said fields is a molecular field.

5. (New) The method of claim 2 wherein each of said fields is selected from the group consisting of: a steric field of a molecule and an electrostatic potential around a molecule.

6. (New) The method of claim 4 wherein at least one of said minimal metric distances is obtained by calculating a maximal overlap between a field on one molecule and a field on another molecule, starting from an orientation of the two fields that is obtained by:

calculating a center of mass and inertia tensor for each molecule; and

translating and rotating one molecule so that its center of mass and at least one inertial axis superimpose respectively with the center of mass and at least one inertial axis of the other molecule.

7. (New) The method of claim 2 wherein each of said fields is a gaussian molecular field.

8. (New) The method of claim 2 wherein each of said fields is an ellipsoidal gaussian representation of a molecule.

9. (New) The method of claim 8 wherein at least one of said ellipsoidal gaussian representations is in sum form.

10. (New) The method of claim 8 wherein at least one of said ellipsoidal gaussian representations is in product form.

11. (New) The method of claim 8 wherein at least one of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing the overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

12. (New) The method of claim 11 wherein said maximizing is calculated by minimizing a value of an ellipsoidal gaussian representation fitness function.

13. (New) The method of claim 12 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

14. (New) The method of claim 2 wherein at least one of said minimal metric distances is obtained by calculating the maximal overlap between a first field on one object and a second field on another object.

15. (New) The method of claim 14 wherein at least one of said minimal metric distances is expressed as a norm of a difference between said first field and said second field.

16. (New) The method of claim 14 wherein at least one of said minimal metric distances is obtained by repeated searches from different starting orientations of said first field.

17. (New) The method of claim 16 wherein each of said repeated searches utilizes a numerical technique selected from the group consisting of: steepest descent, conjugate gradient and Newton-Raphson.

18. (New) The method of claim 17 wherein said numerical technique comprises an analytic derivative of said first field.

19. (New) The method of claim 17 wherein said numerical technique comprises a numerical derivative of said first field.

20. (New) The method of claim 14 wherein at least one of said first field and said second field has associated with it at least one value of an overlap with itself in a different orientation.

21. (New) The method of any one of claims 1, 12, 14 or 17 wherein at least one of said minimal metric distances is calculated on a computer.

22. (New) A method of determining a shape space of a set of molecules, comprising:

choosing an initial set of N molecules;

calculating a distance matrix D wherein each element D_{ij} is a minimal metric distance between a molecule i and a molecule j and wherein said molecule i and said molecule j are in said initial set of molecules;

constructing a metric matrix G from D according to a distance geometry technique;

diagonalizing G, thereby obtaining eigenvalues of G, and obtaining a set of positions in N-1-dimensional space that reproduce the distances in said matrix D to within a tolerance T, wherein each position of said set of positions has N-1 coordinates associated with it;

determining which of the N-1 coordinates that represent positions in shape space of each of the N molecules can be eliminated for every molecule such that a remaining

number, M, of the N-1 coordinates still enables said distance matrix to be reproduced to within said tolerance, T; and

defining the shape space to be an M dimensional subspace occupied by the N molecules.

23. (New) The method of claim 22 wherein said minimal metric distance is calculated as a maximal overlap of a first molecular field of said molecule i and a second molecular field of said molecule j, wherein said first molecular field and said second molecular field are both chosen to be a field selected from the group consisting of steric and electrostatic.

24. (New) The method of claim 22 wherein said minimal metric distance is calculated as a maximal overlap of an ellipsoidal gaussian representation of a molecular field of said molecule i and an ellipsoidal gaussian representation of a molecular field of said molecule j.

25. (New) The method of claim 24 wherein each of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing the overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

26. (New) The method of claim 25 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

27. (New) A method of searching a database of molecules for molecules that are similar to a target molecule, comprising:

determining a shape space of said database of molecules according to the method of claim 22; and

calculating a position of said target molecule in said shape space;

finding a distance in shape space between said target molecule and each of said molecules in said database of molecules, thereby ascertaining which of said molecules is closest in shape space to said target molecule.

28. (New) The method of claim 22 wherein at least one of said minimal metric distances is calculated on a computer.

29. (New) The method of claim 27 wherein the database is stored on a computer.

30. (New) A mathematical method of constructing an ellipsoidal gaussian representation of a molecular field:

choosing a number of ellipsoidal gaussian functions to represent the molecular field, wherein each ellipsoidal gaussian function has a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen starting position and with a randomly chosen starting orientation within the molecular field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

constructing the ellipsoidal gaussian representation from the values of the coordinates of the center, the orientations of the principal axes, the magnitudes of the width factors and the size of the prefactor that give the maximal overlap between the molecular field and the ellipsoidal gaussian functions.

31. (New) The method of claim 30 wherein said maximal overlap is calculated by minimizing a value of an ellipsoidal gaussian representation fitness function.

32. (New) The method of claim 30 additionally comprising the step of expressing the ellipsoidal gaussian representation in sum form.

33. (New) The method of claim 30 additionally comprising the step of expressing the ellipsoidal gaussian representation in product form.

34. (New) The method of claim 30 wherein said number of ellipsoidal gaussian functions is about three or four and said target molecule comprises about twenty atoms excluding hydrogen atoms.

35. (New) The method of claim 30 wherein said number of ellipsoidal gaussian functions is determined by minimizing a fragment adjusted ellipsoidal gaussian representation fitness function wherein said fitness function comprises a sum of the fragment ellipsoidal gaussian representation fitness function and the molecular ellipsoidal gaussian representation fitness function.

36. (New) The method of claim 30 wherein said choosing, centering, forcing and constructing steps are repeated for alternative starting positions and starting orientations of the ellipsoidal gaussian functions so that alternative values of the orientations and the volumes of the ellipsoidal gaussian functions are obtained, corresponding to the overlap between the molecular field and the ellipsoidal gaussian functions obtained from said alternative starting positions and starting orientations.

37. (New) The method of claim 36 wherein values of said alternative orientations and volumes of the ellipsoidal gaussian functions are stored in a computer database.

38. (New) The method of claim 31 wherein said fitness function is an integral over all space of a square of a difference between said molecular field and said ellipsoidal gaussian representation.

39. (New) The method of claim 38 wherein said fitness function is calculated numerically over a lattice of points.

40. (New) The method of claim 38 wherein said fitness function is calculated analytically.

41. (New) The method of claim 38 wherein said molecular field is a steric field represented by a gaussian function centred on each atom.

42. (New) The method of claim 31 wherein said minimizing utilizes a numerical technique selected from the group consisting of: steepest descent, conjugate gradient and Newton-Raphson.

43. (New) The method of claim 42 wherein said numerical technique utilizes an analytic gradient of said molecular field.

44. (New) The method of claim 42 wherein said numerical technique utilizes a numerical derivative of said molecular field.

45. (New) The method of claim 30 wherein said randomly chosen starting position is within an atomic radius of at least one atom of the molecule.

46. (New) The method of claim 30 wherein said forcing is achieved by setting each of said width factors to 1.0.

47. (New) The method of claim 30 wherein said centering additionally comprises orienting said principal axes parallel to x, y and z cartesian axes.

48. (New) The method of any one of claims 30, 31, 36 and 42 wherein said choosing, centering, forcing and constructing steps are carried out on a computer.

49. (New) A method of associating a first atom in a first molecule with a second atom in a second molecule, comprising:

constructing a first ellipsoidal gaussian representation for the first molecule and a second ellipsoidal gaussian representation for the second molecule wherein said first

ellipsoidal gaussian representation comprises a first ellipsoidal gaussian function to which said first atom belongs, and said second ellipsoidal gaussian representation comprises a second ellipsoidal gaussian function to which said second atom belongs;

orienting said first ellipsoidal gaussian function with said second ellipsoidal gaussian function, wherein said first ellipsoidal gaussian function comprises a first center, a first set of width factors and a first set of principal axes and said second ellipsoidal gaussian function comprises a second center, a second set of width factors and a second set of principal axes, so that said first center and said first set of principal axes coincide respectively with said second center and said second set of principal axes and so that the principal axis of said first set of principal axes that corresponds to the largest width factor of said first set of width factors is aligned with the principal axis of said second set of principal axes that corresponds to the largest width factor of said second set of width factors and so that the principal axis of said first set of principal axes that corresponds to the smallest width factor of said first set of width factors is aligned with the principal axis of said second set of principal axes that corresponds to the smallest width factor of said second set of width factors; and

assigning a first atom from said first ellipsoidal gaussian function to the closest atom from said second ellipsoidal gaussian function, thereby associating the first atom in the first molecule with the second atom in the second molecule.

50. (New) The method of claim 49, wherein said orienting is repeated for each of four possible alignments in which said first set of principal axes is parallel to said second set of principal axes.

51. (New) The method of claim 50, wherein said assigning is repeated for each of said four possible alignments and the first atom is associated with the second atom that is closest to it, chosen from each of said four possible alignments.

52. (New) The method of claim 50, wherein said first molecule has a first number of atoms and said second molecule has a second number of atoms that is greater than or equal to said first number of atoms, and wherein said assigning is repeated for each of said four possible alignments and additionally comprises the step of calculating, for each of said four possible alignments, a sum of distances between every atom in the first molecule and the

respective closest atom in the second molecule, and associating the first atom with the second atom corresponding to the alignment for which said sum of distances is smallest.

53. (New) The method of claim 50 wherein each of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing the overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

54. (New) The method of claim 49 wherein, if said first ellipsoidal gaussian representation comprises more than one ellipsoidal gaussian function, said first ellipsoidal gaussian function is identified by a method comprising:

for each of said ellipsoidal gaussian functions, determining a value of an ellipsoidal gaussian representation fitness function between a functional form of said first atom and said ellipsoidal gaussian function; and

identifying said first ellipsoidal gaussian function to be one which has the lowest value of said ellipsoidal gaussian representation fitness function.

55. (New) The method of any one of claims 49 or 54 wherein at least one of said minimal metric distances is calculated on a computer.

56. (New) The method of claim 54 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

57. (New) A method of searching a database for at least one part of a molecule that is similar to at least one part of a target molecular structure wherein the database contains N molecular structures, comprising:

storing in the database an ellipsoidal gaussian representation of a molecular field of each molecular structure in said database wherein each of said ellipsoidal gaussian representations comprises a set of ellipsoidal gaussian functions;

constructing an ellipsoidal gaussian representation of a molecular field of the target molecular structure wherein said ellipsoidal gaussian representation comprises a set of ellipsoidal gaussian functions;

for each of said molecular structures in said database:

calculating a measure of similarity by comparing a subset of said ellipsoidal gaussian functions of a molecular field of the target molecular structure with a subset of said ellipsoidal gaussian functions of a molecular field of said molecular structure in the database; and

reporting to a user a molecule corresponding to said molecular structure in the database if said measure of similarity is greater than a certain specified level.

58. (New) The method of claim 57 wherein the at least one part of a target molecular structure is a vacant space in a pocket on a surface of the target molecular structure.

59. (New) The method of claim 57 wherein said calculating is repeated for each possible subset of said ellipsoidal gaussian functions of a molecular field of the target molecular structure with each possible subset of said ellipsoidal gaussian functions for said molecular structure in the database.

60. (New) The method of claim 57 wherein said molecular field of the target molecular structure and each of said molecular fields of said molecular structures in the database are all selected from the group consisting of: electrostatic potential around a molecule, hydrophobic field, and a steric field of a molecule.

61. (New) The method of claim 57 wherein said measure of similarity is a minimal metric distance.

62. (New) The method of claim 61 wherein said minimal metric distance is obtained by calculating the maximal overlap between a subset of said ellipsoidal gaussian functions of a molecular field of the target molecular structure with a subset of said ellipsoidal gaussian functions of a molecular field of said molecular structure in the database.

63. (New) The method of any one of claims 61 or 62 wherein at least one of said minimal metric distances is calculated on a computer.

64. (New) The method of claim 57 wherein said ellipsoidal gaussian representation of a molecular field of the target molecular structure and each of said ellipsoidal gaussian representations of a molecular field of each molecular structure in said database is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said molecular field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each of said ellipsoidal gaussian functions at a randomly chosen position within said molecular field;

forcing each of said ellipsoidal gaussian functions to initially adopt a spherical shape; and

maximizing the overlap between said molecular field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

65. (New) The method of claim 57 wherein the database is stored on a computer.

66. (New) The method of any one of claims 57, 58 or 64 wherein said constructing is carried out with a computer.

67. (New) The method of claim 57 wherein said storing comprises recording sets of values of the orientations and the volumes of the ellipsoidal gaussian functions that comprise each ellipsoidal gaussian representation in said database and said comparing comprises matching said values with values of the same quantities for said target molecule.

68. (New) The method of claim 67 wherein all possible subsets of said sets of values of the orientations and the volumes of the ellipsoidal gaussian functions are stored in a database in said storing step.

69. (New) A method of organizing a database of N objects for the purpose of facilitating searching the database, comprising:

choosing a first root object at random from said database;

calculating a metric distance from said first root object to each of N-1 other objects in said database;

dividing said N-1 other objects in said database into an upper branch comprising those objects whose distance from said first root object is greater than median, T, of said metric distances and a lower branch comprising those objects whose distance from said first root object is smaller than T;

storing said median with said first root object into a root node data structure associated with said first root object; and

picking a next root object at random from said database;

repeating said calculating, dividing, storing and picking steps for each other object in said database except that said calculating and said dividing steps are performed with just those objects that are on the same branch as said next root object, and except that said first root object is replaced by said next root object, unless a number of objects in said branch has been reduced to one or zero.

70. (New) The method of claim 69 wherein each of said metric distances is a minimal metric distance.

71. (New) The method of claim 70 wherein each of said objects is a field.

72. (New) The method of claim 71 wherein at least one of said minimal metric distances is obtained by calculating a maximal overlap between a field on one object and a field on another object.

73. (New) The method of claim 71 wherein at least one of said minimal metric distances is obtained by repeated searches from different starting orientations of the two fields.

74. (New) The method of claim 71 wherein each of said fields is selected from the group consisting of: a steric field of a molecule and an electrostatic potential around a molecule.

75. (New) The method of claim 74 wherein at least one of said metric distances is a minimal metric distance obtained by calculating a maximal overlap between a field on one molecule and a field on another molecule, starting from an orientation of the two fields that is obtained by:

calculating a center of mass and inertia tensor for each molecule; and

translating and rotating one molecule so that its center of mass and at least one inertial axis superimpose respectively with the center of mass and at least one inertial axis of the other molecule.

76. (New) The method of claim 70 wherein each of said fields is an ellipsoidal gaussian representation of a molecule.

77. (New) The method of claim 76 wherein each of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing the overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

78. (New) The method of any one of claims 69, 75 or 77 wherein the database is stored on a computer.

79. (New) A method of organizing a database of N objects for the purpose of facilitating searching the database, comprising:

selecting K key objects that differ from one another in their respective values of some property and wherein K is less than N;

calculating a minimal distance between each of said K key objects and every other object in said database; and

storing in said database, for each of said K key objects, an ordered list of each database object and its distance from said key object.

80. (New) The method of claim 79 wherein said selecting comprises choosing a representative object from each of K clusters of said objects in said database wherein said clusters are found by a clustering technique.

81. (New) The method of claim 80 wherein said clustering technique is Jarvis-Patrick.

82. (New) The method of claim 79 wherein each of said minimal distances is a minimal metric distance.

83. (New) The method of claim 79 wherein each of said objects is a field.

84. (New) The method of claim 83 wherein at least one of said minimal metric distances is obtained by calculating a maximal overlap between one field and another field.

85. (New) The method of claim 84 wherein at least one of said minimal metric distances is obtained by repeated searches from different starting orientations of the two fields.

86. (New) The method of claim 83 wherein each of said fields is selected from the group consisting of: a steric field of a molecule and an electrostatic potential around a molecule.

87. (New) The method of claim 86 wherein at least one of said minimal metric distances is obtained by calculating the maximal overlap between a field on one molecule and a field on another molecule, starting from an orientation of the two fields that is obtained by:

calculating a center of mass and inertia tensor for each molecule; and

translating and rotating one molecule so that its center of mass and at least one inertial axis superimpose respectively with the center of mass and at least one inertial axis of the other molecule.

88. (New) The method of claim 86 wherein each of said fields is a gaussian molecular field.

89. (New) The method of claim 86 wherein each of said fields is an ellipsoidal gaussian representation of a molecule.

90. (New) The method of claim 89 wherein each of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing the overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of the center, the orientations of the principal axes and the magnitudes of the width factors and the size of the prefactor.

91. (New) The method of claim 79 additionally comprising, before said selecting, calculating a shape space of the database of N objects.

92. (New) The method of claim 91 wherein said K key objects are remote from each other in shape space.

93. (New) The method of Claim 91 wherein said shape space is calculated by a method comprising:

choosing an initial set of N objects;

calculating a distance matrix D wherein each element D_{ij} is a minimal metric distance between object i and object j and wherein said object i and said object j are in said initial set of objects;

constructing a metric matrix G from D according to a distance geometry technique;

diagonalizing G, thereby obtaining eigenvalues of G, and obtaining a set of positions in N-1-dimensional space that reproduce the distances in said matrix D to within a tolerance T, wherein each position of said set of positions has N-1 coordinates associated with it;

removing each of the N-1 coordinates that can be set to zero for every object in said set of objects such that a remaining number, M, of the N-1 coordinates still enables said distance matrix to be reproduced to within said tolerance, T; and

defining the shape space to be an M dimensional subspace occupied by the N objects.

94. (New) The method of any one of claims 79, 87, 90 or 93 wherein at least one of said minimal metric distances is calculated on a computer.

95. (New) The method of claim 79 wherein the database is stored on a computer.

96. (New) The method of claim 89 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

97. (New) A mathematical method of constructing a pseudo-surface of an ellipsoidal gaussian representation of a molecular field, comprising:

for each ellipsoidal gaussian function having three widths whose values are u, v, and w, and having 3 axes A, B and C, in said ellipsoidal gaussian representation:

calculating a volume of said ellipsoidal gaussian function;

obtaining a factor, R, such that a solid ellipsoid with axes whose widths are $R/\sqrt{u}$, $R/\sqrt{v}$ and $R/\sqrt{w}$ has the same volume as said ellipsoidal gaussian function; and

defining a pseudo-surface of said ellipsoidal gaussian function to be a surface of said solid ellipsoid positioned so that its center coincides with the center of said ellipsoidal gaussian function and so that its axis of width $R/\sqrt{u}$ aligns with axis A of said ellipsoidal gaussian function, its axis of width $R/\sqrt{v}$ aligns with axis B of said ellipsoidal gaussian function and its axis of width $R/\sqrt{w}$ aligns with axis C of said ellipsoidal gaussian function.

98. (New) The method of claim 97 wherein the ellipsoidal gaussian representation of the molecular field represents a vacant space in a pocket on a surface of a molecule.

99. (New) The method of claim 97 wherein each of said ellipsoidal gaussian representations is constructed by:

choosing a number of ellipsoidal gaussian functions to represent said field, wherein each ellipsoidal gaussian function comprises a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes;

centering each ellipsoidal gaussian function at a randomly chosen position within said field;

forcing each ellipsoidal gaussian function to initially adopt a spherical shape; and

maximizing an overlap between said field and the ellipsoidal gaussian functions by adjusting the coordinates of each respective center, the orientations of the principal axes, the magnitudes of the width factors and the size of the prefactor.

100. (New) A method of representing a property on a pseudo-surface calculated according to the method of claim 97, comprising:

calculating a value of the property at a number of sample points distributed on the surface of at least one of said solid ellipsoids.

101. (New) A method of comparing values of a property calculated on a first pseudo-surface with values of a property calculated on a second pseudo-surface, wherein said first pseudo-surface and said second pseudo-surface are calculated according to the method of claim 97, comprising:

aligning a first solid ellipsoid in said first pseudo-surface with a second solid ellipsoid in said second pseudo-surface, wherein said first solid ellipsoid comprises a first center, a first set of width factors and a first set of principal axes and said second solid ellipsoid comprises a second center, a second set of width factors and a second set of principal axes, so that said first center and said first set of principal axes coincide respectively with said second center and said second set of principal axes and so that the principal axis of said first set of principal axes that corresponds to a largest width factor of said first set of width factors is aligned with a principal axis of said second set of principal axes that corresponds to a largest width factor of said second set of width factors and so that a principal axis of said first set of principal axes that corresponds to the smallest width factor of said first set of width factors is aligned with a principal axis of said second set of principal axes that corresponds to a smallest width factor of said second set of width factors; and

scaling said first solid ellipsoid to a first sphere of unit radius and said second solid ellipsoid to a second sphere of unit radius; and

comparing, point by point, each of said values of the property on said first sphere to each of said values of the property on said second sphere.

102. (New) The method of any one of claims 97–101 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

103. (New) A method of constructing at least one single ellipsoidal gaussian function representation for a vacant space in a pocket on a surface of a molecule, comprising:

choosing a set of N atoms close to the vacant space;

defining a molecular field function from N spherical ellipsoidal gaussian functions, each of which is centered on one of said N atoms;

placing a test ellipsoidal gaussian function at a random starting point in the vacant space;

minimizing a function of the overlap between said test ellipsoidal gaussian function and said molecular field function; and

repeating said placing and said minimizing for additional random starting points, thereby producing a number of single ellipsoidal gaussian function representations of the vacant space.

104. (New) The method of claim 103 additionally comprising combining two or more of said single ellipsoidal gaussian functions to form an ellipsoidal gaussian representation.

105. (New) The method of claim 103 wherein said minimizing utilizes a numerical technique selected from the group consisting of: steepest descent, conjugate gradient and Newton-Raphson.

106. (New) The method of claim 105 wherein said numerical technique comprises a numerical derivative of said molecular field function.

107. (New) The method of claim 105 wherein said numerical technique comprises an analytic gradient of said molecular field function.

108. (New) A method of searching a database for at least one part of a molecule that fits into a vacant space in a pocket on a surface of a target molecule wherein the database contains N molecular structures, comprising:

storing in the database an ellipsoidal gaussian representation of a molecular field of each molecular structure in said database wherein each of said ellipsoidal gaussian representations comprises a set of ellipsoidal gaussian functions;

constructing an ellipsoidal gaussian representation of a molecular field of the vacant space by combining two or more ellipsoidal gaussian functions obtained by the method of claim 104;

for each of said molecular structures in said database:

calculating a measure of similarity by comparing a subset of said ellipsoidal gaussian functions of a molecular field of the vacant space with a subset of said ellipsoidal gaussian functions of a molecular field of said molecular structure in the database; and

reporting to a user a molecule corresponding to said molecular structure in the database if said measure of similarity is greater than a certain specified level.

109. (New) The method of claim 108 wherein said calculating is repeated for each possible subset of said ellipsoidal gaussian functions of said molecular field of the vacant space with each possible subset of said ellipsoidal gaussian functions for said molecular structure in the database.

110. (New) The method of claim 108 wherein said molecular field of the vacant space and each of said molecular fields of said molecular structures in the database are all selected from a member of the group consisting of: electrostatic potential around a molecule and a steric field of a molecule.

111. (New) The method of claim 108 wherein said measure of similarity is a minimal metric distance.

112. (New) The method of claim 111 wherein said minimal metric distance is obtained by calculating a maximal overlap between a subset of said ellipsoidal gaussian functions of a molecular field of the vacant space with a subset of said ellipsoidal gaussian functions of a molecular field of said molecular structure in the database.

113. (New) The method of claim 108 wherein the database is stored on a computer.

114. (New) The method of claim 108 wherein the vacant space is an active site.

115. (New) The method of claim 108 wherein said molecular field of the vacant space is minus the electrostatic potential in the vacant space.

116. (New) A method of deducing a starting point for the optimization of a fitness function between a molecule and a vacant space in a pocket on a surface of a target molecule comprising:

calculating a maximal overlap between the molecule and the vacant space by the method of claim 112; and

aligning the molecule in the vacant space in the orientation that gives the minimal metric distance between a subset of said ellipsoidal gaussian functions of a molecular field of the vacant space with a subset of said ellipsoidal gaussian functions of a molecular field of said molecular structure in the database.

117. (New) The method of claim 103 wherein said test ellipsoidal gaussian function is initially spherical with a volume set to that of a single carbon atom.

118. (New) The method of claim 103 wherein said test ellipsoidal gaussian function has parameters including a prefactor, three width factors, coordinates of its center, and three mutually orthogonal unit vectors that define the directions of its principal axes and wherein all of said parameters are allowed to vary during said minimizing.

119. (New) The method of claim 103 wherein said set of N atoms consists of all atoms within a specified distance of the vacant space.

120. (New) The method of claim 103 wherein each of said N spherical ellipsoidal gaussian functions has the same volume as the atom upon which it is placed.

121. (New) The method of claim 103 wherein said function of the overlap between said test ellipsoidal gaussian function and said molecular field function comprises a sum of fitting functions between each of said N spherical ellipsoidal gaussian functions, EGF_i, and said test ellipsoidal gaussian function, EGF_T.

122. (New) The method of claim 121 wherein said fitting function is:

$$f = a*V - b*(Q\,(EGF_i,\, V) - (Q\,(EGF_T,\, V))$$

wherein $V = Q\,(EGF_i,\, EGF_T)$ and wherein $Q = \int EGF_i(\bar{r})EGF_T(\bar{r})\mathrm{d}\bar{r}$.

123. (New) The method of claim 104 wherein said pocket is an active site of a protein.

124. (New) A method of representing a property on a surface of an ellipsoidal gaussian representation for an active site, calculated by the method of claim 123, comprising:

for each ellipsoidal gaussian function having three widths whose values are u, v, and w, and having 3 axes A, B and C, in said ellipsoidal gaussian representation:

calculate the volume of said ellipsoidal gaussian function;

obtain a factor, R, such that a solid ellipsoid with axes whose widths are R/√u, R/√v and R/√w has the same volume as said ellipsoidal gaussian function;

define a pseudo-surface of said ellipsoidal gaussian function to be a surface of said solid ellipsoid positioned so that its center coincides with the center of said ellipsoidal gaussian function and so that its axis of width R/√u aligns with axis A of said ellipsoidal gaussian function, its axis of width R/√v aligns with axis B of said ellipsoidal gaussian function and its axis of width R/√w aligns with axis C of said ellipsoidal gaussian function; and

assigning values to points on the surface of said solid ellipsoid by projecting values of a property from said set of N atoms close to said active site.

125. (New) The method of claim 124 wherein said property is a molecular electrostatic potential.

126. (New) The method of claim 103 wherein said pocket is a cleft or is groove-like.

127. (New) The method of claim 103 wherein said molecular field function is in sum form.

128. (New) The method of claim 103 wherein said molecular field function is in product form.

129. (New) The method of claim 103 wherein said molecular field function represents the molecular electrostatic potential in the vacant space.

130. (New) The method of any one of claims 103-105, 121-123 or 126 wherein at least one of said ellipsoidal gaussian representations is calculated on a computer.

131. (New) A method of predicting the biological activity of a molecule of interest, comprising:

for each molecule in a first set of molecules whose biological activities are known, calculating a shape space vector relative to a second set of molecules whose biological activities may or may not be known;

applying a statistical method to said shape space vectors of said first set of molecules to produce a set of weights;

using said weights and a shape space vector of the molecule of interest relative to said second set of molecules to predict a biological activity of said molecule of interest.

132. (New) The method of claim 131 wherein said statistical method is partial least squares.

133. (New) The method of claim 131 wherein said shape space is calculated by a method comprising:

choosing an initial set of N molecules;

calculating a distance matrix D wherein each element D_{ij} is a minimal metric distance between molecule i and molecule j and wherein said molecule i and said molecule j are in said initial set of molecules;

constructing a metric matrix G from D according to a distance geometry technique;

diagonalizing G, thereby obtaining eigenvalues of G, and obtaining a set of positions in N-1-dimensional space that reproduce the distances in said matrix D to within a tolerance T, wherein each position of said set of positions has N-1 coordinates associated with it;

removing each of the N-1 coordinates that can be set to zero for every molecule in said set of molecules such that a remaining number, M, of the N-1 coordinates still enables said distance matrix to be reproduced to within said tolerance, T; and

defining the shape space to be the M dimensional subspace occupied by the N molecules.

134. (New) The method of claim 133 wherein said shape space vector for a molecule in said set of molecules whose biological activities are known is calculated according to:

choosing M+1 sets of coordinates that represent a set that cannot be described at a dimensionality less than M;

calculating distances in shape space between the molecule in set of molecules whose biological activities are known and each of said M+1 sets of coordinates;

generating a set of linear equations for the shape space vector of the molecule in said set of molecules whose biological activities are known, from said distances; and

solving said set of linear equations for said shape space vector.

135. (New) The method of any one of claims 131-134 wherein said set of molecules is stored in a database on a computer.

136. (New) The method of any one of claims 133 or 134 wherein at least one of said minimal metric distances is calculated on a computer.

137. (New) A method of identifying a fragment of a molecule, comprising:

constructing an ellipsoidal gaussian representation of the molecule, wherein said ellipsoidal gaussian representation comprises more than one ellipsoidal gaussian function; and

for each atom in said molecule:

calculating a value of an ellipsoidal gaussian representation fitness function of a functional form of said atom for each of said ellipsoidal gaussian functions;

assigning said atom to an ellipsoidal gaussian representation for which said value is lowest; and

identifying a fragment to be that collection of atoms that is assigned to a particular ellipsoidal gaussian function.

138. (New) A method of assessing the diversity of a set of molecules stored in a database on a computer, the method comprising:

calculating a shape space for the set of molecules in the database; and

defining the diversity of said set of molecules to be a dimensionality of said shape space.

139. (New) The method of claim 138 wherein said shape space is calculated by a method comprising:

choosing an initial set of N molecules;

calculating a distance matrix D wherein each element D_{ij} is a minimal metric distance between molecule i and molecule j and wherein said molecule i and said molecule j are in said initial set of molecules;

constructing a metric matrix G from D according to a distance geometry technique;

diagonalizing G, thereby obtaining eigenvalues of G, and obtaining a set of positions in N-1-dimensional space that reproduce the distances in said matrix D to within a tolerance T, wherein each position of said set of positions has N-1 coordinates associated with it;

removing each of the N-1 coordinates that can be set to zero for every molecule in said set of molecules such that a remaining number, M, of the N-1 coordinates still enables said distance matrix to be reproduced to within said tolerance, T; and

defining the shape space to be the M dimensional subspace occupied by the N molecules.